PRINCIPAL COMPONENTS (PC) AND ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE INTERACTION (AMMI) TREND ANALYSES FOR INCOMPLETE BLOCK AND LATTICE RECTANGLE DESIGNED EXPERIMENTS

BU-1489 -M **June 2000**

Walter T. Federer
Dept. of Biometrics
Cornell University

Russell D. Wolfinger
SAS Institute, Inc.
R52, SAS Campus Dr.
Cary, NC 27513

J. Crossa
Biometrics and Statistics UNit
CIMMYT
Mexico, D.F. Mexico

Keywords: none provided by authors.

Abstract: A principal component is a linear combination of data that has a maximum sum of squares. No other linear combination is associated with a larger sum of squares. Therefore, there analyses could prove useful for describing spatial variation found in field experiments. PC and AMMI analyses have been used in genotype by environment studies. There is a problem with these analyses in that the degrees of freedom for these linear combinations of the data will need to be obtained via simulations. A code for doing this is given elsewhere in this volume. The SAS code allocates a single degree of freedom for each PC but this is not correct. If the F-value associated with a PC is less than the F-value at the 25% level, this PC sum of squares is pooled with the residual sum of squares. Rather than this rule, one may use only a SAS?MIXED procedure and eliminate all effects from the model whose variance components are estimated as zero. The two procedures do not give the same result in general.

BU-1489-M

Title: Principal Components (PC) and Additive Main Effects and Multiplicative Interaction (AMMI) Trend Analyses for Incomplete Block and Lattice Rectangle Designed Experiments.

Authors: W. T. Federer, 434 Warren Hall, Cornell University, Ithaca, NY 14853; e-mail WTF1@cornell.edu, R. D. Wolfinger, SAS Institute R52, SAS Campus Drive, Cary, NC 27513, and J. Crossa, Biometrics and Statistics Unit, CIMMYT, Mexico, D. F. Mexico.

Importance: A principal component is a linear combination of data that has a maximum sum of squares. No other linear combination is associated with a larger sum of squares. Therefore, these analyses could prove useful for describing spatial variation found in field experiments. PC and AMMI analyses have been used in genotype by environment studies. There is a problem with these analyses in that the degrees of freedom for these linear combinations of the data will need to be obtained via simulations. A code for doing this is given elsewhere in this volume. The SAS code allocates a single degree of freedom for each PC but this is not correct. If the F-value associated with a PC is less than the F-value at the 25% level, this PC sum of squares is pooled with the residual sum of squares. Rather than this rule, one may use only a SAS/MIXED procedure and eliminate all effects from the model whose variance components are estimated as zero. The two procedures do not give the same result in general.

References: Federer, W. T., J. Crossa, and J. Franco (1998). Forms of spatial analyses with mixed model effects and exploratory model selection. BU-1406-M, Tech. Report, Dept. of Biometrics, Cornell Univ., Ithaca, NY.
Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. Biometrics 44:705-715.
Moreno-Gonzales, J. and J. Crossa (1998). Combining environments, genotypes, and attribute variables in regression models for predicting the cell-means of multi-environment trials. Theoretical and Applied Genetics 96:803-811.
Zobel, R. W. (1990). A powerful statistical tool for understanding genotype-by-environment interaction. In *Genotype-byEnvironment Interaction and Plant Breeding* (Editor: M. S. Kang) Louisiana State University, Baton Rouge , pp. 126-140.

Data: The example used to illustrate the SAS code is the balanced lattice square designed experiment given in Table 12.5 of Cochran and Cox (1957). *Experimental Designs*. Wiley, NY. By altering the program appropriately, the code can also be used for incomplete block and row-column designed experiments. Also, PCs can be computed using the correlation matrix or the variance-covariance matrix. The SAS default uses the correlation matrix.

```
/* Here the data the data are included but an infile statement may be used to input data. */
data original;
  input y rep row col grad treat; /* grad is the linear regression coefficient on column order. */
  label row='incomplete block';
  datalines;
9.0   1 1 1 -3 10
20.3  1 1 2 -1 12
17.7  1 1 3  1  9
26.3  1 1 4  3 11
 4.7  1 2 1 -3  2
 9.0  1 2 2 -1  4
 7.3  1 2 3  1  1
 8.3  1 2 4  3  3
 9.0  1 3 1 -3 14
 6.7  1 3 2 -1 16
11.7  1 3 3  1 13
 4.3  1 3 4  3 15
 4.0  1 4 1 -3  6
 5.0  1 4 2 -1  8
 5.7  1 4 3  1  5
```

14.3	1	4	4	3	7
19.0	2	1	1	-3	5
8.7	2	1	2	-1	12
13.0	2	1	3	1	15
15.7	2	1	4	3	2
12.0	2	2	1	-3	10
6.0	2	2	2	-1	7
15.3	2	2	3	1	4
12.0	2	2	4	3	13
12.7	2	3	1	-3	16
6.3	2	3	2	-1	1
1.7	2	3	3	1	6
13.0	2	3	4	3	11
3.7	2	4	1	-3	3
3.7	2	4	2	-1	14
8.0	2	4	3	1	9
13.3	2	4	4	3	8
17.0	3	1	1	-3	10
7.0	3	1	2	-1	15
10.3	3	1	3	1	8
1.3	3	1	4	3	1
11.3	3	2	1	-3	9
12.3	3	2	2	-1	16
3.0	3	2	3	1	7
5.3	3	2	4	3	2
12.3	3	3	1	-3	12
8.7	3	3	2	-1	13
8.0	3	3	3	1	6
9.3	3	3	4	3	3
30.3	3	4	1	-3	11
22.3	3	4	2	-1	14
11.0	3	4	3	1	5
12.7	3	4	4	3	4
5.0	4	1	1	-3	16
10.3	4	1	2	-1	12
5.7	4	1	3	1	8
12.7	4	1	4	3	4
2.7	4	2	1	-3	11
6.7	4	2	2	-1	15
10.3	4	2	3	1	3
5.7	4	2	4	3	7
1.0	4	3	1	-3	1
10.3	4	3	2	-1	5
11.3	4	3	3	1	9
11.7	4	3	4	3	13
11.0	4	4	1	-3	6
19.0	4	4	2	-1	2
20.7	4	4	3	1	14
29.7	4	4	4	3	10
2.0	5	1	1	-3	3
5.0	5	1	2	-1	16
4.0	5	1	3	1	5
13.7	5	1	4	3	10
9.3	5	2	1	-3	6
1.7	5	2	2	-1	9
6.3	5	2	3	1	4

```
12.3 5 2 4  3 15
16.7 5 3 1 -3 12
 4.3  5 3 2 -1   7
18.7 5 3 3  1 14
 8.7  5 3 4  3   1
16.7 5 4 1 -3 13
30.0 5 4 2 -1   2
25.7 5 4 3  1 11
14.0 5 4 4  3   8
run;

/* principal component analysis.  */
proc sort data=original;
  by  rep col row;
proc transpose data=original prefix=row out=origr(drop=_name_);
  by  rep col;
  var y;
proc princomp data=origr prefix=rpc out=rowvar noprint;
  by  rep;
  var row1-row4; /* Four rows in the design. */
proc sort data=original;
  by  rep row col;
proc transpose data=original prefix=col out=origc(drop=_name_);
  by  rep row;
  var y;
proc princomp data=origc prefix=cpc  out=colvar noprint;
  by  rep;
  var col1-col4; /* Four columns in the design. */

/* expand data sets and merge */
data cc;
  set colvar;
  array colv{4} col1-col4;
  do col = 1 to 4;
    y = colv{col};
    output;
  end;
  drop col1-col4;
data rr;
  set rowvar;
  array rowv{4} row1-row4;
  do row = 1 to 4;
    y = rowv{row};
    output;
  end;
  drop row1-row4;
proc sort data=rr;
  by  rep row col;
data ana;
  merge original cc rr;
  by  rep row col;

/* analysis of variance using the principal components, non-nested */
proc glm data=ana;
  class rep treat;
  model y=rep treat cpc1 cpc2 rpc1 rpc2 cpc1*rpc1 cpc1*rpc2
```

```
    cpc2*rpc1 cpc2*rpc2 ;
run;

/* using the principal components, nested */
proc glm data=ana;
  class rep treat;
  model y=rep treat cpc1(rep) cpc2(rep) rpc1(rep) cpc1*rpc1(rep) cpc1*rpc2(rep)
  cpc2*rpc1(rep) cpc2*rpc2(rep) ;
run;

/* For random blocking effects, use the following code. */
proc mixed data = ana;
  class rep treat row col;
  model y = treat;
  random rep cpc1(rep) cpc2(rep) rpc1(rep) cpc1*rpc1(rep) cpc1*rpc2(rep)
  cpc2*rpc1(rep) cpc2*rpc2(rep);
  lsmeans treat;
run;

/* Using the textbook analysis of the design as in Cochran and Cox (1957), page 493. */
proc glm data=ana;
  class rep row col treat;
  model y=rep treat row(rep) col(rep);
run;

/* For an AMMI trend analysis, PC within row within replicate, use the following code. */
proc glm data = ana;
  class rep treat row col;
  model y = rep treat row(rep) rpc1*row(rep);
run;
proc mixed data = ana;
  class rep treat row col;
  model y = treat;
  random rep row(rep) rpc1*row(rep);
  lsmeans treat;
run;
```

An abbreviated form of the output from this program is given below.

```
Dependent Variable: Y
                               Sum of          Mean
Source              DF         Squares         Square         F Value    Pr > F
Model               27         2723.926541     100.886168       5.93      0.0001
Error               52          884.611459      17.011759
Corrected Total     79         3608.538000

          R-Square         C.V.        Root MSE            Y Mean
          0.754856         37.82239     4.124531            10.90500

Dependent Variable: Y

Source              DF        Type I SS     Mean Square    F Value       Pr > F
REP                  4        31.563000        7.890750       0.46       0.7619
TREAT               15      1244.202000       82.946800       4.88       0.0001
CPC1                 1       937.165951      937.165951      55.09       0.0001
CPC2                 1        28.780965       28.780965       1.69       0.1991
```

RPC1	1	465.277940	465.277940	27.35	0.0001
RPC2	1	7.694324	7.694324	0.45	0.5042
CPC1*RPC1	1	1.186538	1.186538	0.07	0.7927
CPC1*RPC2	1	7.642424	7.642424	0.45	0.5057
CPC2*RPC1	1	0.325597	0.325597	0.02	0.8905
CPC2*RPC2	1	0.087801	0.087801	0.01	0.9430

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	4	31.5630000	7.8907500	0.46	0.7619
TREAT	15	371.7330707	24.7822047	1.46	0.1571
CPC1	1	913.7729632	913.7729632	53.71	0.0001
CPC2	1	75.9641822	75.9641822	4.47	0.0394
RPC1	1	466.8813930	466.8813930	27.44	0.0001
RPC2	1	8.7776829	8.7776829	0.52	0.4758
CPC1*RPC1	1	1.8655460	1.8655460	0.11	0.7419
CPC1*RPC2	1	7.7122489	7.7122489	0.45	0.5037
CPC2*RPC1	1	0.3476938	0.3476938	0.02	0.8869
CPC2*RPC2	1	0.0878007	0.0878007	0.01	0.9430

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	54	3503.080701	64.871865	15.38	0.0001
Error	25	105.457299	4.218292		
Corrected Total	79	3608.538000			

R-Square	C.V.	Root MSE	Y Mean
0.970776	18.83400	2.053848	10.90500

General Linear Models Procedure

Dependent Variable: Y

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	4	31.563000	7.890750	1.87	0.1470
TREAT	15	1244.202000	82.946800	19.66	0.0001
CPC1(REP)	5	1034.531257	206.906251	49.05	0.0001
CPC2(REP)	5	46.108785	9.221757	2.19	0.0879
RPC1(REP)	5	480.204652	96.040930	22.77	0.0001
CPC1*RPC1(REP)	5	262.322233	52.464447	12.44	0.0001
CPC1*RPC2(REP)	5	71.654226	14.330845	3.40	0.0177
CPC2*RPC1(REP)	5	117.938209	23.587642	5.59	0.0014
CPC2*RPC2(REP)	5	214.556340	42.911268	10.17	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	4	31.563000	7.890750	1.87	0.1470
TREAT	15	84.006127	5.600408	1.33	0.2576
CPC1(REP)	5	1145.198035	229.039607	54.30	0.0001
CPC2(REP)	5	56.156010	11.231202	2.66	0.0462
RPC1(REP)	5	401.390987	80.278197	19.03	0.0001
CPC1*RPC1(REP)	5	197.143931	39.428786	9.35	0.0001
CPC1*RPC2(REP)	5	67.960365	13.592073	3.22	0.0222
CPC2*RPC1(REP)	5	138.518519	27.703704	6.57	0.0005
CPC2*RPC2(REP)	5	214.556340	42.911268	10.17	0.0001

The MIXED Procedure

Least Squares Means

Effect	TREAT	LSMEAN	Std Error	DF	t	Pr > |t|
TREAT	1	7.54548158	1.72555784	29	4.37	0.0001
TREAT	2	10.37683195	1.75725118	29	5.91	0.0001
TREAT	3	9.37048456	1.56493910	29	5.99	0.0001
TREAT	4	12.34206915	1.57505850	29	7.84	0.0001
TREAT	5	11.45961200	1.61696971	29	7.09	0.0001
TREAT	6	9.11047013	1.69326123	29	5.38	0.0001
TREAT	7	7.86631476	1.75974765	29	4.47	0.0001
TREAT	8	11.11913803	1.67931870	29	6.62	0.0001
TREAT	9	12.18761611	1.64388978	29	7.41	0.0001
TREAT	10	14.08160555	1.92299755	29	7.32	0.0001
TREAT	11	13.13295404	1.92869896	29	6.81	0.0001
TREAT	12	11.40375822	1.67381535	29	6.81	0.0001
TREAT	13	10.59779026	1.54426804	29	6.86	0.0001
TREAT	14	12.13599166	1.70542148	29	7.12	0.0001
TREAT	15	9.18610407	1.62649961	29	5.65	0.0001
TREAT	16	12.56377795	1.63547347	29	7.68	0.0001

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	49	2928.370083	59.762655	2.64	0.0029
Error	30	680.167917	22.672264		
Corrected Total	79	3608.538000			

R-Square	C.V.	Root MSE	Y Mean
0.811511	43.66382	4.761540	10.90500

Dependent Variable: Y

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	4	31.563000	7.890750	0.35	0.8433
TREAT	15	1244.202000	82.946800	3.66	0.0012
ROW(REP)	15	1093.015500	72.867700	3.21	0.0032
COL(REP)	15	559.589583	37.305972	1.65	0.1197

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	4	31.563000	7.890750	0.35	0.8433
TREAT	15	319.452083	21.296806	0.94	0.5350
ROW(REP)	15	1026.755833	68.450389	3.02	0.0049
COL(REP)	15	559.589583	37.305972	1.65	0.1197

... ...

Effect	TREAT	LSMEAN	Std Error	DF	t	Pr > |t|
TREAT	1	6.37311825	2.38292075	25	2.67	0.0130
TREAT	2	10.61497350	2.15761108	25	4.92	0.0001
TREAT	3	7.93436160	2.06059560	25	3.85	0.0007
TREAT	4	12.28827249	1.99475730	25	6.16	0.0001
TREAT	5	10.76991952	1.97476997	25	5.45	0.0001
TREAT	6	8.24325882	2.17029859	25	3.80	0.0008
TREAT	7	5.71709277	2.38681041	25	2.40	0.0244
TREAT	8	10.24396596	2.09650099	25	4.89	0.0001
TREAT	9	12.25845651	2.11969951	25	5.78	0.0001
TREAT	10	14.74535577	2.35708086	25	6.26	0.0001

TREAT	11	15.30490826	2.42463576	25	6.31	0.0001
TREAT	12	13.57092699	2.14533617	25	6.33	0.0001
TREAT	13	11.45714987	2.01082282	25	5.70	0.0001
TREAT	14	13.94971794	2.10372824	25	6.63	0.0001
TREAT	15	8.62267164	2.09619203	25	4.11	0.0004
TREAT	16	12.38585010	2.14469670	25	5.78	0.0001